

# Genetic Classification and Language Contact

Michael Noonan

*University of Wisconsin-Milwaukee*

## 1. INTRODUCTION<sup>1</sup>

Until recently, within orthodox linguistic circles, there probably would have been little to say about the relation between genetic classification of languages and language contact except to say that the latter was irrelevant to the former. Languages might indeed come into contact and various aspects of grammar might be borrowed from one language to another, but such borrowing did not affect a language's genetic classification, which was determined by the retention of inherited morphemes through the process of regular generational transmission, and which was scientifically established by comparing only inherited, not borrowed, morphemes through the comparative method.

Over the years, a number of linguists have expressed reservations about many of the assumptions underlying this view, and in this paper I will examine both the orthodox view and various alternatives to it. The paper will begin with a discussion of what it might mean to say that two languages are genetically related. I will follow this with a discussion of models of genetic relatedness, paying special attention to the widely-accepted family tree model and the assumptions that underlie it. I will then consider various outcomes of language contact and discuss what sorts of models of genetic relatedness these are most compatible with. Lastly, I will address the topic of speciation – the creation of new languages – and language contact.

## 2. WHAT DO WE MEAN BY THE GENETIC CLASSIFICATION OF LANGUAGES?

To the lay person, it is perhaps not immediately obvious what could be meant by the 'genetic' classification of languages. Languages are not living organisms for which descent, implying as it does birth, parenting, and death, could unproblematically apply. A layperson's guess might be that, in speaking about the genetics of languages, linguists are equating the genetics of language with the genetics of the people who speak them.

From the standpoint of contemporary linguistics, this is surely not the case. Regardless of what one thinks about the various theories which have been put forward asserting a strong genetic influence on the structuring of languages, all mainstream linguists would agree that any child can learn any language natively, and that the child's genetic background has no effect on the genetic classification of the language the child learns.<sup>2</sup> It is also surely the case that mainstream linguists would agree that languages are in important ways cultural artifacts, consisting in some sense of cultural *memes* (Dawkins 1976), and that these linguistic memes are subject to the same sorts of pressures and changes that other memic systems are subject to.

So, languages are not linked genetically to the people who speak them, and they are asserted to be in some sense cultural artifacts: a specific language is learned by a child because of the circumstances of his or her birth and rearing, and languages, like other aspects of culture, change through time. When speaking of other cultural artifacts, we can, of course, talk informally about 'descent', in the sense that one can say that in some ways American culture descends from British culture. But it is certainly not standard practice in such discussions to draw family trees and speak about genetic relatedness. Why is language different?

---

<sup>1</sup> I would like to thank Edith Moravcsik for helpful discussions about the issues discussed in this paper.

<sup>2</sup> The contemporary disassociation of the genetics of people and the genetics of the languages they speak did not characterize the early 19<sup>th</sup> century Romantics who created the family tree model which dominates modern conceptions of genetic relationships among languages. For them, there was a straightforward connection between culture, language, and people (Thom 1995). It was only later that tracing the origins of people and tracing the origins of languages came to be considered separate subjects of inquiry, though interestingly there has been renewed interest in this connection in recent times, for example in Cavalli-Sforza *et al* 1988, Cavalli-Sforza 2000, Ruhlen 1994.

One reason language is different is that, unlike something as amorphous as a culture, languages have traditionally been viewed as consisting of a finite set of relatively easily identifiable entities – words, grammatical affixes, rules, etc. – which are organized systematically and which can in principle be compared in a straightforward manner. Such comparisons can form the basis for assessing relatedness among languages.

Linguistic constructs, of course, can be related in a number of ways. The field of linguistic typology, for example, is concerned with an assessment and evaluation of the similarities and differences of various linguistic features or combinations of features across languages. But typological classifications are not genetic classifications.<sup>3</sup> The traditional genetic classification of a language, at least at the higher taxonomic levels, tells us very little about the structure of the language – less, for example, than knowing where in the world a language is spoken.<sup>4</sup> Traditional genetic classification of a language also tells us very little about the source of morphemes employed by speakers: the morpheme inventories of many languages can easily be shown to contain a majority of forms borrowed from languages outside their immediate taxonomic units. Chantyal, as just one example, is classified as a Sino-Tibetan language, yet its morpheme inventory overwhelmingly consists of borrowings from Indo-European Nepali (Noonan 2003); English is classed as a Germanic language, though its morpheme inventory is largely drawn from Italic and Greek.

If the genetics of speakers, typological similarity, and a percentage assessment of the source of morphemes in a language are not relevant for telling us what genetic relationships among languages mean, then what is? Three approaches to this issue can be found in the literature, though they are usually merely assumed rather than explicitly argued for.

The first approach I will label the *generational transmission* approach. This is conceptually the simplest and is almost never discussed explicitly but is merely assumed. In this way of looking at things, assessing the genetic relatedness of languages amounts to assessing the history of the generational transmission of linguistic traditions. By ‘generational transmission of linguistic traditions’ I mean the acquisition by children of essentially the same linguistic system that their parents acquired as children.<sup>5</sup> In this way, English is a West Germanic language because if one traces the history of the generational tradition of the language we now call English, we will find that it merges with the linguistic traditions we call Dutch, German, etc. approximately 1400 years ago. Similarly, Irish and Hindi, despite their radically different typologies, can be shown through various sorts of evidence to be traceable back through generational transmission to a common language called Proto-Indo-European.

The second of these approaches I will label the *essentialist* approach, following Croft (2000:197). This position maintains that there are certain linguistic features, consisting both of grammatical morphemes and characteristic morpho-syntactic features, that must be transmitted along a genetic line for a language to be considered a member of a given taxonomic unit. This is not to say that these features over time cannot change. It maintains only that in assessing potential mother/daughter rela-

---

<sup>3</sup> As Robins (1990:186f) points out, the 18<sup>th</sup> century view of genetic relations among languages was largely typological in the modern sense. The *Encyclopédistes* did not consider French a descendant of Latin because the grammars were so different; instead, they considered French to be a continuation of Gaulish, which had adopted a Latin vocabulary. Similarly, Sir William Jones, who famously described a kinship relation between Sanskrit, Latin, and Greek, did not regard Hindi as a descendant of Sanskrit because of their very different grammars.

<sup>4</sup> Knowing that Irish and Hindi are Indo-European languages tells us almost nothing about the structures of the two languages; knowing that Hindi is South Asian tells us a good deal more.

<sup>5</sup> In assessing genetic relatedness, it is not important that *all* instances of generational transmission follow this model since a language can acquire new adult speakers. It is necessary only that there be *some* instances of unbroken transmission in the linguistic community.

tionships, these features must be transmitted; language relatedness is assessed along chains of transmission of these features from mother language to daughter language. Croft attributes this position to Thomason & Kaufman (1988), who assume it as part of their discussion of normal *vs* abnormal transmission of language, but something similar seems to have been accepted by other scholars over a long period as we will see in the discussion of the treatment of creoles and other language varieties in genetic linguistics.

The first two approaches will ordinarily yield the same analyses, but they differ conceptually, and this conceptual difference has consequences in certain cases, as will be discussed in below in §4. Both approaches are fully compatible with the *comparative method*, a technique for verifying genetic relations that was developed in the 19<sup>th</sup> century and reached its mature form under the Neo-grammarians. While both approaches are compatible with the comparative method, the essentialist approach incorporates some of its assumptions into the approach itself. It does this by asserting that some of the material utilized by the comparative method to *demonstrate* linguistic relatedness is required to *have* a parent/offspring genetic relationship at all.<sup>6</sup>

The third approach, which I will label the *hybrid* approach, is really a conflation of several distinct approaches, ranging from the wave theorists (*e.g.* Schmidt 1872) to contemporary comparativists [*e.g.* the Sino-Tibetanists Benedict (1972), Chappell (2001), Matisoff (2001), Pulleyblank (1998)], to contemporary theoreticians (Croft 2000, Laks 2002), and to creolists (Holm 1988, Mufwene 2002). What these approaches have in common is the idea that, at least in some circumstances, languages may be mixed, hybrids of otherwise valid taxonomic units.<sup>7</sup> This is different from the generational transmission approach and the essentialist approach, both of which disallow hybridity in the assessment of genetic relations. The hybrid approach is not fully compatible with the comparative method, which requires descent along a single genetic line for its operation, though many of those who have subscribed to hybrid approaches use the comparative method, but assert that it is not always applicable – or not straightforwardly applicable – in all cases.<sup>8</sup>

A hybrid approach takes the position that a language is a collection of entities (morphemes, grammatical constructs, etc.) that may have multiple sources. At some point, the mixture of forms may become so great as to preclude the assignment of the language to a specific taxon within a hierarchy of taxonomic levels, though it might still be easily placed within a higher level.<sup>9</sup> Most linguists these days would concede that true ‘mixed languages’ exist, *e.g.* Copper Island Aleut, Michif, Media Lengua, etc., but would relegate them to a category outside the normal development of languages – that is, outside any genetic line. Others would include creoles in the category of hybrid languages, while still others would include in this category at least some non-creoles as well.

---

<sup>6</sup> It’s worth emphasizing that the comparative method is not in itself a model of genetic relatedness, but rather a technique for demonstrating it. It’s not the only such technique currently in use: for example, Greenberg and his associates (Greenberg 2005, Ruhlen 1994), controversially, have used a technique at odds with the comparative method for assessing genetic relationships. See Campbell (1997, 2003) for critiques of this approach and other contemporary alternatives to the comparative method.

<sup>7</sup> Some of the scholars listed are less than clear about whether they believe that *in principle* there are hybrids or whether under some circumstances it is impossible to say what taxonomic unit a language belongs to. Others explicitly or implicitly allow hybridity, *e.g.* Benedict, Croft, Laks, and Holm.

<sup>8</sup> “I remain doubtful, however, of the possibility of successfully reconstructing Chinese linguistic history strictly from the evidence of modern dialects by the traditional comparative method as applied to languages without a written tradition. A major difficulty is that the Stammbaum or branching-tree model that is implied by the traditional comparative method is totally unrealistic in the case of Chinese.” (Pulleyblank 1998:200)

<sup>9</sup> So, for example, it might not be possible to classify Cantonese within a taxon at the level of other Chinese languages, but it can still be placed within Sinitic.

In the discussion that follows, I will have rather less to say about hybrid approaches than the generational transmission and essentialist approaches simply because the latter two have been accepted by many more linguists over the years and because none of the hybrid approaches has yet attracted a substantial and influential number of adherents.

Before proceeding further, it might be worth asking what genetic classification is good for. It has already been stated that genetic classification is not always useful in providing information about the structure of a language or its morpheme inventory, the more so the higher up the taxonomic ladder one goes. Information about where in the world a language is spoken provides more useful information about grammatical structure, but we don't have classifications of languages by region that are comparable to genetic classifications. On the positive side, however, genetic classification has proven a boon to historical linguistics, providing the superstructure around which theories of language change have developed over the last two centuries. Such classifications also, potentially, provide information of considerable historical value. Typologists use genetic classifications to explain similarities among languages and as a consideration in constructing crosslinguistic samples. And, of course, most of us find satisfying the classification of familiar things: typically the first thing a linguist will ask on being told of an unfamiliar language is: "What family does it belong to?"<sup>10</sup>

### 3. MODELS OF LANGUAGE FAMILIES IN GENETIC LINGUISTICS

In the last section, I discussed three approaches to the question of what genetic relatedness for languages might mean. These approaches are primarily concerned with language creation and not directly with how more remote relations among languages might be dealt with. It remains now to discuss models of more remote relations.

The term 'genetic' strongly suggests a biological model or analog for the classification of languages, and indeed linguists have employed biologically-inspired models.<sup>11</sup> It is worth noting at the outset of this section that an explanation for why a biological analogy employing the term 'genetic' should be applied to languages alone among cultural artifacts is virtually absent in the linguistics literature,<sup>12</sup> where by long tradition going back over two hundred years most linguists have simply assumed the validity of a biological analog in linguistic classification.<sup>13</sup>

Within a biologically-inspired framework, there are at least two possible classes of interpretations of genetic relatedness. One could conceive of languages as *unitary organisms* and consider relatedness in a way analogous to that of individual animals or plants, which can be related via lineages

---

<sup>10</sup> In fact, genetic classifications of languages have served many purposes within and outside linguistics, including various intellectual and even political causes. For example, awareness of the classification of Rumanian as a Romance language ultimately affected its writing system [from cyrillic to roman] and encouraged the movement to purge the language of Slavic and other non-Italic elements.

<sup>11</sup> The relationship between linguistics and biology was not simply a one-way expropriation of ideas. Nettle (1999:4) points out that the success of Indo-Europeanists in the 19<sup>th</sup> century in charting the histories of languages encouraged Darwin in his development of the theory of evolution.

<sup>12</sup> Croft (2000) and Mufwene (2001, 2007) are notable exceptions.

<sup>13</sup> In the 19<sup>th</sup> century, the doctrine of 'biological naturalism' in linguistics did make explicit the connection between languages and living organisms. Ivić (1965), in her history of linguistics, says the following about August Schleicher, an early and very influential comparativist [p44]: "His [Schleicher's] method grew out of his conception that language was a living organism, independent of man, whose line of development was determined by the general biological laws of evolution: a language is born, lives for a certain time, gives life to another, younger language which in time replaces it, in its turn to be continued by one of its own offshoots; this language, like man, has a "genealogical tree", i.e. a common ancestor from which numerous related progeny have developed as branches of the tree (hence Schleicher's theory is called the theory of biological naturalism in linguistics, and is known under the name of the "Stammbaum", or "pedigree" theory)."

created through sexual or asexual reproduction. Alternatively, one could conceive of a language as a *population*, either of speakers or of linguistic constructs<sup>14</sup>, or even of a population of speakers each with his/her idiolect [and hence his/her own set of linguistic constructs].<sup>15</sup> Population models of this sort might adopt a species analogy for understanding genetic relatedness.<sup>16</sup>

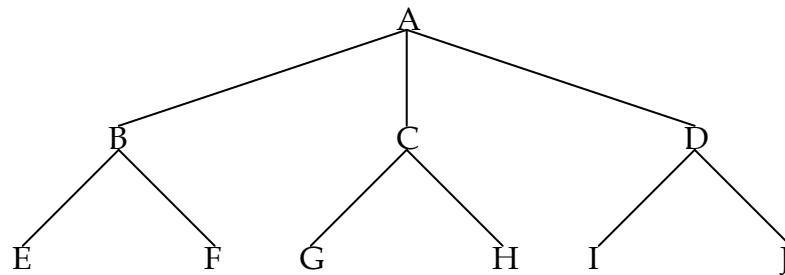
In linguistics, the *unitary organism* model was the one adopted by historical linguists in the early 19<sup>th</sup> century; this model has survived as the received mode of understanding genetic relations to the present day. Within this model, two languages are said to be genetically related if they descend from a common ancestor. Since it is at least possible that all languages descend from a common ancestor, languages are usually claimed to be related only if their relatedness can be established through the comparative method or some alternative procedure.<sup>17</sup>

In principle, a unitary organism model could adopt either an asexual [parthenogenetic] or a sexual model for conceptualizing genetic relatedness. The established model, known as the *family tree* [or *Stammbaum*] model, adopted parthenogenetic [asexual] reproduction as the mode for understanding genetic relationships among languages. The expressions *mother/ancestor language* and *daughter language* are components of the model and reflect the original analogy, as do the notions of *language birth* and *language death*.

Of the approaches to the nature of genetic classification discussed in §2, the generational transmission approach and the essentialist approach are fully compatible with the family tree model and for the most part seem to presuppose it, though in a few special cases they can allow for developments that are disallowed by the family tree model. Hybrid approaches are compatible with all the alternative models, namely a unitary organism model that supposes [or allows] sexual reproduction, and the various sorts of population models as a few scholars (*e.g.* Croft 2000, Mufwene 2001, 2007) have made explicit.

**3.1 THE FAMILY TREE MODEL:** The family tree model assumes that any set of related languages descends from a single ancestor according to the parthenogenetic model of biology. Within this model, lineages may be represented by means of diagrams like (1) where A represents the immediate ancestor of B, C, and D and the common ancestor of E, F, G, H, I, and J, all of which can be said to be related.

(1)



The classical family tree model assumes, following this parthenogenetic analogy, that there can be no special genetic relationship between, say, F and G other than their common descent from A. The model also assumes that influence on a language, even massive influence, cannot affect its genetic status, any more than external influence on a bacterium could, in older biological models, affect its status within its lineage. There is no linguistic feature or set of features which determine the genetic

<sup>14</sup> By *linguistic construct* I mean any unit or construction within a language.

<sup>15</sup> Croft (2000), for example, views languages as populations of utterances produced by a set of communicating speakers.

<sup>16</sup> Mufwene (2001) argues that languages should be conceptualized in a way analogous to *species*.

<sup>17</sup> See footnote 6 and the references cited there for discussion. It suffices here to say that while the comparative method and the various alternative approaches differ in how relationships may be established, they suppose an identical model of how languages may be related. It is the latter issue that concerns us here.

status of a language; rather, it is the circumstance of its birth that determines this. In this way, anything that is borrowed from another language does not affect its genetic status. Further, the model supposes that splits [the birth of new languages] are always final and produce independent linguistic systems.

This last point, that splits produce independent linguistic systems, is an important component of the model and conforms to the parthenogenesis analogy. Within this model, a language is treated as an entity analogous to, for example, a bacterium in a line of parthenogenetic descent. It is divisible in the sense that it may, asexually, give 'birth' to new languages, but it cannot 'merge' with another language, it cannot engage in sexual reproduction [there is always a single, unique ancestor for any lineage], and it is not composed of 'parts' that may merge or split in ways not consistent with the model generally. This last proviso concerns the status of dialects: their status within the model is exactly like that of languages. The model in (1) could diagram the relations of dialects within a language as well as a set of related languages.

The conceptual simplicity of the family tree classification schema follows, in part, from a set of basic assumptions which, taken together, make it possible, even necessary, to reject completely the effects of language contact in assessing genetic relationships. Indeed, one of the problems one often encounters in establishing genetic relationships according to the family tree model is the problem of stripping away the effects of contact so as to reveal the core of 'native' material necessary for the comparative method. Contact is thus irrelevant for the determination of genetic relationships with the comparative method, though the effects of contact can prove an obstacle to its implementation.<sup>18</sup>

This aspect of the family tree model follows from a literal interpretation the unitary organism cum parthenogenetic reproduction analogy. A language is composed of many linguistic constructs, but there is no threshold beyond which a language ceases to be a member of a lineage due to change of these linguistic constructs. That is, it is possible for a language to change all the constructs inherited from a remote ancestor and still stay within a lineage. For example, if in (1) above, language E could be demonstrated to have descended from B, and B could be demonstrated to have descended from A, then it would follow that E is a descendent of A and within the family of languages defined by A even if A and E share no more linguistic constructs than any randomly selected pair of languages might share. So, membership within a lineage is not dependent on the possession of any particular feature or set of features, or even on the possession of any shared feature or set of features. It is based simply on the fact of common descent, no matter how this is determined. Nonetheless, in the usual course of things, common descent implies a certain number of shared features with other members of a lineage, and these common features are required by the comparative method for establishing membership within the lineage.

It's important to emphasize here that acceptance of the family tree model of genetic relationship does not in itself preclude an appreciation of the role of language contact in the historical development of languages. Instead, what is implied by the model is that contact, along with other modes of language change, is irrelevant for genetic classification. Nonetheless, issues relating to contact situations have formed the bases for criticisms of the model. We will consider several classes of such criticisms below.

In sum, the family tree model of genetic relationships rests on a set of assumptions that can be summarized below, all of which, in one way or another, follow from the unitary organism and parthenogenesis analogies:

- (2) a. languages are unitary systems: they are wholes, not entities defined by their parts [*the unitary organism analogy*]

---

<sup>18</sup> Campbell (2004:212) notes that "there is no provision in the comparative method for dealing directly with borrowings."

- b. two languages are genetically related if they descend from a single, common ancestor [*the parthenogenesis analogy*]
- c. new languages can only be created by splitting off from an existing language [*the parthenogenesis analogy*]
- d. linguistic splits are final and produce independent linguistic systems [*the parthenogenesis analogy*]
- e. no linguistic feature or set of features is required for genetic relationships to exist between two languages (though such features are required for *establishing* such relations)<sup>19</sup> [*the unitary organism analogy*]
- f. language contact is irrelevant for determining genetic relationships [*the unitary organism and parthenogenesis analogies*]

(2e) probably requires some additional comment. Shared features are required for the operation of the comparative method, but the comparative method is not in itself a component of the family tree model, but rather a methodology traditionally allied with it. As noted, the methodology for establishing genetic relationships utilized by Greenberg and his associates (Greenberg 2005, Ruhlen 1994) is at odds with the comparative method, but is quite compatible with the family tree model.

3.2. ALTERNATIVES TO THE FAMILY TREE MODEL: The alternative models implied by the biological analogy have not received much attention, at least until recently. A unitary organism model employing sexual reproduction has been notably absent from discussions of genetic relatedness, although Mufwene in a series of publications (e.g. 2001, 2007) has held that speciation [language splits] in the evolution of a language often come about via language contact, which could suggest a sort of sexual model of speciation, at least in some instances. Croft (2000) also discusses a sexual analogy in creation of mixed languages. Population models have been explored by various linguists in recent times [again by Croft (2000) and Mufwene (2001, 2007)], though the full consequences of models of this sort for language relatedness and our conceptualization of language generally have yet to be fully explored. In a population model, the gene pool, or its analog, could be considered variable, and new genetic material may be acquired by the species through hybridization, as well as by mutation and other means compatible with contemporary biological models.

It should be noted that the models of genetic relatedness discussed above represent conceptually the simplest sorts of models: those based on the simplest analogies with the biological domain. One could, of course, propose more complex models. For example, many plants can reproduce both sexually and asexually and a model of genetic relatedness for languages could be based on the possibility of both sorts of reproduction, which might include principled reasons for deciding which sort of reproduction has taken place in any given instance. Croft (2000) notes this possibility.

The one clear advantage of the family tree model over all the biologically-inspired alternatives is that, at a macroscopic level, it provides a conceptually simple description of what happens to families of languages, such as Indo-European, during the course of their evolution and where individual languages should be placed within the set of their known relatives. This simple classificatory system has undeniable appeal, and it is notably the case that linguists have attempted to replicate the apparent early success of Indo-European and Semitic linguistics in establishing family trees for all the other proposed language families.

#### 4. GENETIC CLASSIFICATION AND LANGUAGE CONTACT

---

<sup>19</sup> The essentialist approach to genetic relatedness discussed in §2 claims only that certain features must be preserved in immediate mother-daughter relationships, but not in more remote relationships.

In the sections that follow, I will discuss a number of situations involving language contact and see how these might be interpreted to affect genetic classification.

**4.1 BORROWING IN THE ABSENCE OF SPECIATION:** I will begin the survey with a discussion of a few basic instances of borrowing in the absence of speciation [birth of new languages] and where generational transmission of a linguistic tradition is not disrupted. The central issue here is what effect borrowing under these conditions has on the genetic classification of languages.

Borrowing resulting in the transfer of vocabulary items, or even longterm bilingual situations resulting in the transfer of syntactic constructions from one language to another, would have no effect on the genetic classification of a language if we agree that genetic classification is simply a record of the history of normal generational transmission of a linguistic tradition. In principle, this would be true even if the borrowing were massive and even if the language underwent metatypic change – that is, if the language changed from one morpho-syntactic type to another.

If we assume the essentialist approach, borrowing would have an effect on genetic classification only in the event of large-scale borrowing associated with the creation of new languages, *i.e.* in instances of speciation. Such instances will be discussed below. Where speciation is not involved, the essentialist approach allows for even massive borrowing without affecting genetic classification: in such cases, it is in complete agreement with the generational transmission approach.

Most hybrid approaches also consider borrowing to affect genetic classification only in the event of speciation. It's difficult to find instances where contemporary scholars have claimed that a language changed genetic classification through borrowing without having undergone speciation, but some individual cases might be understood in this light. For example, Wexler (1991) has claimed that Eastern Yiddish is relexified Judeo-Sorbian; given that Yiddish is generally thought to be a Germanic language, this would appear to amount to a claim that Yiddish changed its genetic status without speciation, although Wexler's (1991) specific claim is that Yiddish is *really* a Slavic language despite relexification and that Western and Eastern Yiddish are genetically unrelated, the former being a Germanic language. Wexler (2002), however, claims explicitly that Modern Hebrew is relexified Yiddish, and that the two languages are 'genetically related' [p3], but the creation of Modern Hebrew should probably be seen as a case of speciation, and hence it would fall outside the category of borrowing in the absence of speciation. This situation is discussed further in §4.2.

Most instances of languages that have borrowed so heavily that their genetic status would be somehow in dispute would probably qualify as mixed languages and also to be the product of speciation. One possible partial exception is Dongolawi, described by Heine & Kuteva (2001). In the course of its evolution, Dongolawi, a Nubian language, came under the influence of Nobiin, another Nubian language, though in a different branch of the family. The language retained much of its native vocabulary, but borrowed most of its grammar, including grammatical morphemes, from Nobiin. This seems not to have been a case of speciation, but rather of evolutionary change within a tradition of generational transmission. Heine & Kuteva claim [p401] that modern Dongolawi is a 'daughter' of both pre-contact Dongolawi and Nobiin: certainly the elements that the comparative method would use to establish the genetic affiliation for this language are mixed. The language can be viewed as a mixed language, but does not seem to have undergone speciation into its mixed form. On the face of it, Heine & Kuteva's claim that modern Dongolawi is a daughter of two languages is not consistent with the generational transmission or essentialist approaches, but rather supposes some sort of hybrid model.

**4.2 SUBSTRATIC INFLUENCE:** It's useful to separate out substratic influence from the cases of borrowing discussed in the last section. By substratic influence I mean a situation whereby a language previ-



ously spoken by a community affects the language the community later comes to speak. For example, the English spoken in Ireland has been affected in a variety of ways by Irish, the language previously spoken by the population of the country. It has often been asserted that French is the product of Vulgar Latin with a Gaulish and probably Vasconic substratum.

The reason for separating substratic influence from other instances of borrowing is that with substratic influence we have situations in which generational transmission of linguistic traditions is disrupted; we may also have speciation, though this is not necessarily the norm. In this section, I will discuss substratic influence in a general way; the effects of substratic influence in the languages traditionally designated as creoles will be discussed separately in §4.4.

Instances of substratic influence are not problematic for the generational transmission or essentialist approaches as long as there are some members of the community who continue the generational transmission of the linguistic tradition, and, in the case of the essentialist approach, some 'core' elements are included in the language of the new speakers.

Nonetheless, we can imagine situations involving substrata that would be challenging for these two approaches, in particular the generational transmission approach. For example, suppose that an entire community decided to adopt a new language in the absence of any native speakers of that language. What would be the genetic affiliation of this new linguistic variety? We have a specific instance that fits this scenario in the case of Modern Hebrew. It was noted in the last section that Wexler (2002) has claimed that Yiddish and Modern Hebrew are genetically related since Yiddish formed the substratum for a revived Hebrew. Most scholars place Modern Hebrew in the Semitic family, and indeed the classical comparative method would unproblematically treat the language this way. Nonetheless, unbroken generational transmission did not take place in such a way that would link Modern Hebrew with the other Semitic languages. The essentialist approach, however, can be interpreted to fit the connection between Modern Hebrew and Classical Hebrew, and so the classification of Modern Hebrew as Semitic is basically essentialist.

**4.3 KOINÉIZATION AND THE LOSS OF AUTONOMY:** Related linguistic varieties frequently come into contact, for instance in national institutional settings, as a result of migration or trade, and in colonial 'tabula rasa' situations [*i.e.* where there were no varieties of the language spoken in the region before]. One result of such contact may be the creation of a *koiné*. The original koiné [the Koiné] was a variety of Ancient Greek which had come to supplant other, local Greek dialects during the Hellenistic and Roman periods. All Modern Greek dialects, save one, are descendants of the Koiné. The Koiné was based mostly on the Athenian dialect, but included many elements from other dialects and involved a certain amount of simplification: the disappearance of irregularities in favor of structurally regular forms.

The term koiné has come to be used for any variety which supplants heteronomous varieties and serves as a means of intercommunication between speakers of these varieties. This comes about as a result of dialect leveling, *i.e.* the loss of distinctive features in favor of features with a high degree of mutual intelligibility and/or high prestige. Sometimes this involves a fair amount of dialect mixture, though this needn't be the case. Where dialect mixture is involved, the process of creating the koiné can be referred to as *koinéization*. Koinéization has probably been a fairly common feature of the history of languages. For example, it seems to have operated at least twice in the history of the mainland Scandinavian varieties (Dahl 2001). The new koiné may exist alongside all or some of the varieties that existed before koinéization, or it may supplant them completely, either regionally or everywhere the varieties were spoken.

Koinéization is not problematic for the approaches using the family tree model (generational transmission and essentialist approaches) as long as it can be maintained that the koiné is essentially

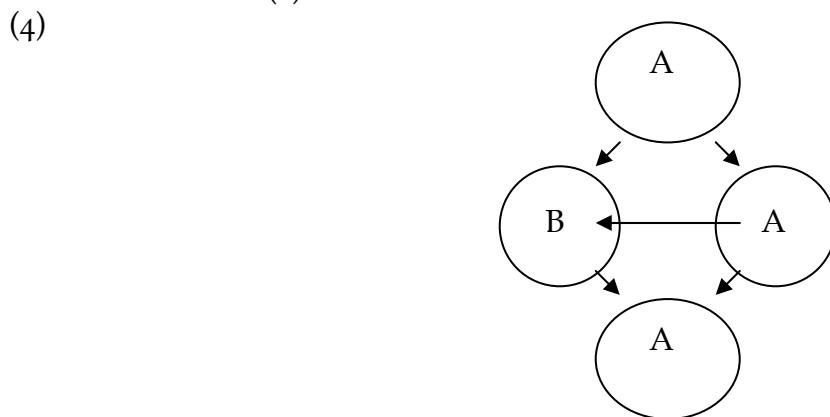
a continuation of one of the original varieties, as in (3a) below, where the koiné is a continuation of D and only the koiné survives. (3b), however, diagrams a scenario where the koiné cannot be non-arbitrarily placed under any of the previously existing varieties because it incorporates too many features from more than one variety. Cases like this would be consistent with some hybrid approaches, but not with the family tree model. The question is, do we find real examples that are like (3b)?



There do appear to be cases that fit the model of (3b). Trudgill (2004) discusses instances of dialect creation in colonial tabula rasa situations where it would be impossible in a non-arbitrary way to assign the koiné resulting from the mixture of many dialect forms to any source dialect. A more interesting, if more unusual case, is the Romansch variety known as Rumantsch Grischun. This variety was created artificially by the Swiss linguist Heinrich Schmidt, who applied a statistical approach to the forms found in the surviving Romansch dialects. It was not intended to supplant the dialects, but rather to be used where there is a need for a single variety intelligible to all (Haiman & Benincà 1992). In that role, it has achieved a far amount of success and is now widely used in publications and official signs.

Rumantsch Grischen is not obviously a descendant of any of the traditional dialects – it was deliberately designed not to be. It would appear, therefore, to fit the schema diagrammed in (3b), except that it did not displace the other varieties.

A related kind of situation is discussed by Dixon (1997). Okinawan had achieved the status of an independent language after 700 years of independent evolution from Japanese, and was the official language of the Ryukyu Kingdom. It has since remerged with Japanese after the formal annexation of the Ryukyu Kingdom by Japan in 1879. Speakers of the various Ryukyu dialects now consider their varieties to be dialects of Japanese, indicating that the language, which had achieved autonomy (in the sense of Trudgill 2000) as a state language, is now heteronomous with Japanese after many decades of intense pressure from the Japanese authorities. The loss of autonomy has been accompanied by a degree of linguistic convergence with Japanese. The situation as described would suggest a model like that in (4):



After the split, which produces an independent B, A and B come into contact again, with the result that A strongly influences B, and B loses its status and is incorporated back into A. In the case of Okinawan, A represents Japanese and B the various Ryukyu varieties, including Okinawan. Situations like this have probably been fairly frequent in the history of languages. For example, the Gallo-

Romance varieties that produced French, Gascon, and Provençal have remerged back into French through a process like that described for Japanese and Okinawan.

Situations that can be characterized by diagram (4) are not compatible with the family tree model, which supposes that linguistic splits are final and cannot be undone. On the other hand, neither the generational transmission nor the essentialist approaches are incompatible with (4). This is one of the few cases where these approaches and the family tree model make different predictions about possible developments.

**4.4 CREOLES:** Creoles are usually defined as languages which develop from pidgins when the latter take on native speakers. In the early stages of its development, a creole typically has a lexifier language, a language which is the source of the great bulk of its vocabulary. The grammar of creoles is usually thought to be simpler than that of its lexifier: in fact some, like McWhorter (2001), claim to be able recognize creoles by their radically simple structure alone.<sup>20</sup>

In a long tradition dating back to the early 19<sup>th</sup> century, creoles have typically been excluded from the family trees of their lexifiers. If one accepts the characterization of creoles that is enshrined in most introductions to linguistics and summarized in the paragraph above, the exclusion of, say, Jamaican Creole from the Germanic languages and Haitian Creole from the Romance languages follows from assumptions about the nature of genetic relatedness embodied in the two main approaches.

Given the characterization above, the generational transmission approach would exclude creoles from the genetic lines of their lexifiers because these languages are not the products of regular generational transmission of a linguistic tradition given their origin in pidgins: a pidgin is, by definition, not a native language, so parents could not be transmitting to their children the linguistic tradition that they themselves acquired as children. The essentialist approach would exclude them because crucial core grammatical features are missing from the radically simplified structure of creoles because in the historical progression from lexifier language to pidgin to creole the grammatical *essence* of the lexifier language has been lost.

Over the last few years, a number of scholars have challenged the narrative about the genesis of creoles summarized in the first paragraph of this section.<sup>21</sup> The new, revisionist narrative is the product of a line of research into the histories of creoles and the populations that speak them. The revisionist position is based on the idea that the creoles of the Atlantic and the Indian Oceans, the prototypes for this class of languages, did not develop from pidgins. Instead, most of these creoles initially developed via normal generational transmission among communities which included significant numbers of European native speakers (albeit mostly of non-standard varieties) along with other peoples whose composition varied from place to place. That is, in origin, these varieties were essentially no different from colonial varieties generally. Where they came to differ from other colonial varieties, such as Brazilian Portuguese, Quebecois French, North American English, etc., has to do with subsequent history. For example, the proto-creoles in regions that experienced the rise of large-scale plantation culture came to be spoken by large numbers of new immigrants, whose languages formed substrata which influenced the subsequent development of the languages, and since these new immigrants were mostly slaves, who experienced increasing segregation from other native

---

<sup>20</sup> An entire issue of *Linguistic Typology* (Volume 5-2/3, 2001) was devoted to the issue of whether creoles constitute a special, and especially simple, structural type.

<sup>21</sup> See, for example, Mufwene (2003, 2007), Chaudenson (2001), and Ansaldo *et al* (2007). My discussion is based largely on Mufwene.

speakers of the colonial languages, the proto-creoles were socially isolated and developed along different paths from other varieties of the language.

The sketch presented in the last paragraph is a simplification of the revisionist view, and the reader is encouraged to consult the references provided.<sup>22</sup> However, given our limited objectives here, it will suffice and it remains to be seen how acceptance of the revisionist model would affect views of the genetic status of these languages. In the generational transmission approach, creoles would now be seen as legitimate offspring of metropolitan languages: Jamaican Creole is a Germanic language and Haitian Creole is a Romance language. This would follow because these languages were the product of regular generation transmission by at least some components of the speaker base at each stage in their development.

Within the essentialist approach, creoles would also now be seen as descendants of the metropolitan languages. While many grammatical features of the standard versions of the metropolitan languages may be absent, their absence was not abrupt and resulted from internal changes, perhaps spurred on by substratic influence and even influence from other creole varieties.

**4.5 MIXED LANGUAGES:** Mixed languages have attracted a good deal of attention over the last two decades.<sup>23</sup> Versteegh (2007) points out that there may not be a linguistically valid category of mixed languages, but I will assume here that there is for the sake of discussing how such languages might be dealt with from the standpoint of genetic linguistics.

In some respects, mixed languages are related to mixed code varieties found in bilingual situations. The primary difference is that mixed languages have achieved autonomy as linguistic systems and some degree of stability in the sense that the variability found in mixed code varieties is considerably reduced. It has been claimed (*e.g.* Winford 2003, Dixon 1997) that mixed languages arise only under a specific set of circumstances, namely where bilingual communities feel the need for a distinctive, in-group language and create a stable mixed-code variety for this purpose. Languages like Michif and Copper Island Aleut fit this model. In these special mixed languages, there is little or no simplification of the two components of the language because the population producing the new system is fully competent in both.

The generational transmission approach runs into the interesting problem that, if the population is bilingual from childhood, as the originators of Michif and Copper Island Aleut probably were, then both systems are transmitted normally, and the mixed language, being a product of this normal generational transmission, has in a real sense two genetic parent languages. The essentialist model, which looks to the transmission of core grammatical forms, would find, in the case of Michif, that the NP is fully French and the VP is fully Cree: the essence of these systems seems to have been transmitted, but only in a component of the grammar. Mixed languages have no place in the family tree model. Hybrid approaches, needless to say, would find such situations less paradoxical.

**5. LANGUAGE CONTACT AND SPECIATION:** Speciation in the sense used here, *i.e.* the creation of new languages, is a complex issue in several respects. One complication is the language-dialect problem in cases of dialect continua: at what point should historic varieties of a language be considered separate languages? Needless to say, mutual intelligibility is a factor, but hardly a criterial one: there are many instances of languages with mutually unintelligible dialects. Similarly, varieties that are mutu-

---

<sup>22</sup> The authors of these works suggest that there may not be a linguistically significant category of creole languages, and what the languages so labeled have in common is their development during a particular period of world history and the fact that they are spoken by people of non-European descent.

<sup>23</sup> Useful information and analysis can be found in Thomason & Kaufman (1988), Bakker & Mous (1994), Winford (2003), and Mous (2003).

ally intelligible may be considered separate languages. In the end, socio-cultural attitudes are the determining factors in such cases.

The problem which concerns us here is the degree to which language contact may result in speciation. Language contact can induce change in the language of various sorts, through simple mechanisms like borrowing, but also through more complex ones resulting from extensive bilingualism and/or substratic influence. Change arising from any or all of these mechanisms may have unequal consequences for different varieties of a language, resulting in reductions in mutual comprehensibility. This will increase the likelihood of speciation, though in itself it does not cause it. In the end, speciation remains largely a matter of social attitudes and purely linguistic considerations are only a factor, though not a negligible one.

6. FINAL THOUGHTS: How one sees the consequences of language contact affecting genetic relations depends on one's adherence to particular approaches to the nature of genetic relationships and models of remote relations. These are bound to evolve as linguists become more familiar with the histories of the various sorts of contact situations. We already see the emergence of new sorts of models in the works of Croft and Mufwene. We can expect to see further developments along those lines.

### References

- Aikhenvald, Aleksandra & R.M.W. Dixon, eds. 2001. *Areal Diffusion and Genetic Inheritance*. Cambridge: Cambridge University Press.
- Ansaldo, Umberto & Stephen Matthews, Lisa Linn, eds. 2007. *Deconstructing Creole*. Amsterdam: John Benjamins.
- Bakker, Peter & Maarten Mous, eds. 1994. *Mixed languages. 15 case studies in language intertwining*. Amsterdam: IFOTT.
- Benedict, Paul. 1972. *Sino-Tibetan: a Conspectus*. James A. Matisoff, contributing editor. New York: Cambridge University Press.
- Campbell, Lyle. 1997. *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press.
- Campbell, Lyle. 2003. 'Beyond the Comparative Method?' In Barry Blake and Kate Burridge, eds. *Historical linguistics 2003*. p33-58. Amsterdam: John Benjamins.
- Campbell, Lyle. 2004. *Historical Linguistics: An Introduction*. Cambridge: MIT Press.
- Cavalli-Sforza, L.L. & A. Piazza, P. Menozzi, J. Mountain. 1988. 'Reconstruction of human evolution: bringing together genetic, archeological, and linguistic data.' *Proceedings of the National Academy of Sciences of the USA*, 85.6002-6.
- Cavalli-Sforza, Luigi Luca. 2000. *Genes, Peoples, & Languages*. New York: North Point Press.
- Chappell, Hilary. 2001. 'Contact and diffusion in Sinitic.' In Aikhenvald & Dixon (2001), p328-357.
- Chaudenson, Robert. 2001. *Creolization of Language and Culture*. London: Routledge.
- Croft, William. 2000. *Explaining Language Change*. Harlow: Longman.
- Dahl, Östen. 2001. 'The origin of the Scandinavian languages.' In Östen Dahl & Maria Koptjevskaja-Tamm, eds. *The Circum-Baltic Languages: Typology and Contact*. Amsterdam: John Benjamins. P215-235.
- Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Dixon, R.M.W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Greenberg, Joseph. 2005. *Genetic Linguistics: Essays on Theory and Method*, edited by William Croft. Oxford: Oxford University Press.
- Haiman, John & Paola Benincà. 1992. *The Rhaeto-Romance Languages*. London: Routledge.

- Heine, Bernd & Tania Kuteva. 2001. 'Convergence and divergence in Africa.' In Aikhenvald & Dixon (2001), p392-411.
- Holm, John. 1988. *Pidgins and Creoles I: Theory and Structure*. Cambridge: Cambridge University Press.
- Ivić, Milka. 1965. *Trends in Linguistics*. Amsterdam: Mouton.
- Laks, Bernard. 2002. 'Le comparatisme: de la généalogie à la génétique.' *Langages*, 146:19-45.
- Matisoff, James. 2001. 'Genetic versus contact relationship: prosodic diffusibility in South-East Asian languages.' In Aikhenvald & Dixon (2001), p291-327.
- McWhorter, John. 2001. 'The world's simplest grammars are creole grammars.' *Linguistic Typology* 5-2/3:125-166.
- Mous, Maartin. 2003. *The Making of a Mixed Language: The Case of Ma'a/Mbugu*. Amsterdam: John Benjamins.
- Mufwene, Salikoko. 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Mufwene, Salikoko. 2002. 'Competition and selection in language evolution.' *Selection* 3.45-56.
- Mufwene, Salikoko. 2007. 'Population movements and contacts: competition, selection, and language evolution.' *Journal of Language Contact* 1:63-91.
- Nettle, Daniel. 1999. *Linguistic Diversity*. Oxford: Oxford University Press.
- Noonan, Michael. 2003. 'Chantyal.' In Randy LaPolla and Graham Thurgood, eds. *The Sino-Tibetan Languages*. London: Routledge. p315-335.
- Pulleyblank, Edwin G. 1998. 'Qieyun and Yunjing: The Essential Foundation for Chinese Historical Linguistics.' *The Journal of the American Oriental Society*, Vol. 118:2, p200-216.
- Robins, R.H. 1990. *A Short History of Linguistics*. 3<sup>rd</sup> ed. London: Longman.
- Ruhlen, Merritt. 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Thom, Martin. 1995. *Republics, Nations and Tribes*. London: Verso.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Trudgill, Peter. 2000. *Sociolinguistics: An Introduction to Language and Society*. London: Penguin Books.
- Trudgill, Peter. 2004. *New Dialect Formation: The Inevitability of Colonial Englishes*. Oxford: Oxford University Press.
- Versteegh, Kees. 2007. 'Review of Mous: *The Making of a Mixed Language: The Case of Ma'a/Mbugu*. *Studies in Language* 31-4:891-900.
- Wexler, Paul. 1991. Yiddish – the fifteenth Slavic language. A study of partial language shift from Judeo-Sorbian to German. *International Journal of the Sociology of Language* 91.9-150.
- Wexler, Paul. 2002. *Two-tiered Relexification in Yiddish: Jews, Sorbs, Khazars, and the Kiev-Polessian Dialect*. Berlin: Mouton de Gruyter.
- Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.